Introducing the GEneva Music-Induced Affect Checklist (GEMIAC): A Brief Instrument for the Rapid Assessment of Musically Induced Emotions

Eduardo Coutinho

University of Liverpool

Klaus R. Scherer

University of Geneva

Eduardo Coutinho, Department of Music, University of Liverpool, United Kingdom; Klaus R. Scherer, Swiss Center for Affective Sciences, University of Geneva, Switzerland.

This research was supported by an ERC Advanced Grant in the European Community's 7th Framework Programme under grant agreement 230331-PROPEREMO (Production and perception of emotion: an affective sciences approach) to Klaus Scherer and by the National Center of Competence in Research (NCCR) Affective Sciences financed by the Swiss National Science Foundation (51NF40-104897) and hosted by the University of Geneva. We are very thankful to Christelle Gillioz, Cristina Soriano, and Johnny Fontaine for the lengthy discussions and insightful suggestions at various stages of the development of this checklist.

Correspondence concerning this article should be addressed to Eduardo Coutinho, University of Liverpool, Department of Music, 80-82 Bedford Street South, Liverpool L69 7WW, United Kingdom. Phone: +44 151 794 2527; E-mail: e.coutinho@liverpool.ac.uk

Word count: 7189

Abstract

The systematic study of music-induced emotions requires standardized measurement instruments to reliably assess the nature of affective reactions to music, which tend to go beyond garden-variety basic emotions. We describe the development and conceptual validation of a checklist for rapid assessment of music-induced affect, designed to extend and complement the Geneva Emotional Music Scale. The checklist contains a selection of affect and emotion categories that are frequently used in the literature to refer to emotional reactions to music. The development of the checklist focused on an empirical investigation of the semantic structure of the relevant terms, combined with fuzzy classes based on a series of hierarchical cluster analyses. Two versions of the checklist for assessing the intensity and frequency of affective responses to music are proposed.

Keywords: Emotion, Feeling, Music, Measurement, Checklist

Introducing the GEneva Music-Induced Affect Checklist (GEMIAC): A Brief Instrument for the Rapid Assessment of Musically Induced Emotions

A critical issue in experimental studies of musically induced emotions is the development of appropriate methods for measuring and classifying emotional states. In many studies, participants are requested to listen to music and then to describe their emotional reaction to the music by rating predetermined affect terms. Most researchers in this area use ad hoc lists of emotion terms or borrow emotion models and measures from nonmusical areas of emotion research. Given that the qualities of the emotions most commonly felt while listening to different types of music tend to differ from those experienced in other contexts (e.g., Zentner, Grandjean, & Scherer, 2008), it can be argued that a domain-specific approach should be used. In other words, empirical research on emotion induction by music should identify terms referring to emotions and other affective feelings that are frequently used in the specific context of listening to music. Therefore, the development of an appropriate affect taxonomy that permits capturing a more realistic and informative spectrum of musical emotions and attempt their quantification is of paramount importance for future research in this area.

Although there were some early attempts to develop rating scales specific to music-induced emotions (e.g., Hevner, 1936), only recently instruments systematically derived from expressions directly linked to musical experiences have been proposed (Asmus, 1985; Bartel, 1992; Juslin & Laukka, 2004; Zentner, Grandjean, & Scherer, 2008). In particular, the Geneva Emotional Music Scale (GEMS; Zentner, Grandjean, & Scherer, 2008) is based on a music-specific model of emotion, and is currently the most systematic instrument available to assess the nature and structural organization of musically induced emotions. The GEMS was developed on the basis of extensive empirical work that aimed to identify which labels listeners

of different kinds of music choose to describe their emotional experiences. The scale consists of 45 emotion terms, representing nine emotion factors (hereafter "GEMS subscales" or "GEMS factors")—Wonder, Transcendence, Tenderness, Nostalgia, Peacefulness, Power, Joyful activation, Tension, and Sadness—describing a wide range of feelings of emotion experienced while listening to music. The scale is widely used and, since its release, has been cited in more than 541 publications.

While the research conducted to develop GEMS has arguably produced the most comprehensive survey of felt emotions while listening to music developed to date, it does not encompass the whole gamut of relevant musical emotions, particularly with respect to negative valence. Most likely, this is a consequence of emotions felt while listening to music being more frequently positive (e.g., Juslin & Laukka, 2004; Sloboda & O'Neill, 2001; Zentner et al., 2008), and thus the most frequently reported feelings in the surveys conducted by Zentner at al. (2008) as the basis for scale development. Another potential limitation of the GEMS is the relatively small number of musical genres used in scale construction (Classical, Jazz, Rock, Pop and World). In addition, the final GEMS scales were fine-tuned for Classical (western art) music (see Zentner et al., 2008, Study 4). In sum, it is plausible to assume that negative feelings are somewhat underrepresented in the GEMS and it is an open question as to how well the GEMS can assess emotional experiences associated with a wide range of musical genres. In fact, the impetus for the development of a more comprehensive checklist came from our experiences in piloting a short version of the GEMS in intensive interactions with composers, interpreters, and members of the audience at concerts of contemporary art music (for example at the Arcana 2010 festival; http://www.arcanafestival.at). Many of the respondents indicated that some of their most important emotional reactions to the different varieties of contemporary music represented at the festival were not covered by the GEMS items (e.g., "being bored",

"enjoying pure beauty", "being full of enthusiasm"). It was also mentioned that in many cases, music does not produce strong emotions but rather subtle, almost fleeting, affective states without a definite quality (such as "being moved").

Furthermore, the relatively large number of items in the GEMS, requires a considerable investment of time by the raters. While this is psychometrically desirable, it limits the use of the instrument in field studies designed to assess immediate emotional responses to music in specific listening contexts such as concert halls, churches, or home settings, which require a more parsimonious, rapid assessment instrument. To accommodate these needs, Zentner and colleagues proposed a short (GEMS-25) and an ultra-short (GEMS-9) version of the GEMS (see Zentner & Eerola, 2010, p. 206). These short scales were created by (a) reducing the number of items per GEMS factor on the basis of confirmatory factor analysis (GEMS-25); and (2) using the factor labels for the nine primary GEMS factors illustrated with three emotion adjectives each (GEMS-9). While the GEMS-9 is a more suitable instrument for rapid assessment of musically induced emotions, it emphasizes the overarching factor labels rather than the fuzzy set of meaning facets provided by the different terms constituting the respective factor. In fact, some researchers using the GEMS-9 scale (Aljanaki, Wiering, & Veltkamp, 2016; Torres-Eliard, Labbe, & Grandjean, 2011; Vuoskoski & Eerola, 2010) found that listeners have difficulties in interpreting some of the factors (Wonder and Transcendence), which leads to inconsistent ratings (Vuoskoski & Eerola, 2010), a less frequent use of some scales (Aljanaki, Wiering, & Veltkamp, 2016), and consequently to a decrement of measurement quality.

These observations highlight three important issues in relation to the application of the GEMS in empirical studies. First, some important feeling classes, particularly those with negative valence, are currently lacking in the scale. Second, the short

version of the GEMS (the GEMS-9) is limited in terms of interpretation since using the factor labels only sacrifices the richness of meaning of the various terms that define each factor. Third, the Wonder and Transcendence scales are somewhat unclear for listener raters.

A new checklist for the rapid assessment of music induced affect

In order to address the above issues, and to achieve a compromise between the need for a brief checklist for use in concert settings and the desire to retain the variety and richness of the verbal descriptions of affective musical experiences, we decided to extend and complement the GEMS (as originally developed in our laboratory; Zentner et al., 2008) with a brief checklist for the rapid assessment of affective feelings elicited by music - the GEneva Music-Induced Affect Checklist (GEMIAC). The aim of the development was 1) to preserve the well-established nine dimensions of the GEMS, 2) to add a number of affective reaction categories considered to be highly relevant by researchers, composers, musicians, and listeners (including negative feelings), and 3) to use a fuzzy set approach with semantically overlapping yet distinct terms to represent the richness and complexity of meaning of the major categories or factors (rather than factor labels).

Design

As a first step in the construction of the GEMIAC, we selected a subset of terms belonging to each of the nine GEMS factors (see Zentner et al., 2008, Table A1, p. 519), since these terms were based on extensive inductive development and therefore provide an ideal starting point to build the new checklist. This overlap allows cross-referencing of data obtained with the original GEMS. Based on the comments from raters collected in several empirical studies with the GEMS, we then added terms for a small set of new classes of feeling terms to enlarge the spectrum of musical emotions to be captured by the new instrument.

In order to remedy the problem of assigning a single verbal label to describe each factor, we propose using fuzzy feeling classes, that is, clusters of terms defining a fuzzy overarching concept or feeling category (feeling being defined here as the mental representation of the different components of an emotion episode; Scherer, 2009, pp. 1318-1324; see also Damasio & Carvalho, 2013). To keep the complexity manageable and to assure homogeneity, we decided to define each category by two affective terms that were closely related without being completely overlapping synonyms. In consequence, our goal was to find the pairs of terms that best describe each GEMS factor and each new proposed feeling class by empirically analyzing the semantic relationship between terms and clearly identifying defined feeling clusters (i.e., pairs of terms that tend to cluster together empirically).

In order to examine the empirical structure of the list of selected terms in order to identify the underlying feeling classes, we ran a number of pilot studies to examine the semantic overlap and the dimensionality of the items contained in our preliminary list. Specifically, we obtained judgments of dissimilarity between the feelings of emotion described by a set of terms depicting affective experiences during music listening. These data were then analyzed by means of hierarchical clustering methods to reveal natural groupings between the various terms on the basis of their affective meaning. The goal was to determine the best pairs of terms to describe each of the GEMS subscales, as well as to evaluate the inclusion of new subscales. The further development of the final checklist described below was based on the results of the initial pilot studies.

Selection of affective terms and fuzzy feeling classes

We first selected terms that best represent each of the original nine GEMS factors to form nine 2-term feeling classes. We identified those terms from the original word list used to construct the GEMS that are sufficiently similar to form a coherent fuzzy

concept, yet as different as possible from other pairs or terms forming a particular feeling class. We also used lexicostatistic inventories to select terms that are frequently used in the English language. As a consequence, rather than confining ourselves to the terms included in the final version of the GEMS reported by Zentner and colleagues (2008, Table A1, p. 519; hereafter referred to as "GEMS-A1"), our choice of terms had to be expanded. The reasons for this involved concerns about the meaning of certain terms, the conceptual distance between terms associated with the same GEMS factor, and the translation into English of some of the terms (the GEMS was developed in French and later translated to English).

In consequence, as described in greater detail below, some of the terms in the final version of the GEMS were reformulated or replaced by new terms. However, virtually all of these new terms had been part of the analyses conducted by Zentner et al. (2008, Table 2, p. 504; hereafter "GEMS-T2"), which shows the percentage of listeners who reported having felt various affective states somewhat or a lot. These terms had not been chosen during the various factors analyses that led to the final version of GEMS. Other terms in our new selection correspond to the labels used by Zentner et al. (2008) to describe the factors. Although our ultimate goal was to create a new fuzzy-set checklist with only two items per feeling class, for development purposes, in some cases we selected more than two items from each of the nine GEMS subscales to empirically determine which pairs of terms would produce the most coherent subscales. The full list of items thus selected is shown in Table 1, and the detailed choices and the justification for our selection of terms to represent the original nine GEMS subscales in the new rapid checklist are described in Supplemental Material Section A.

-- Insert Table 1 here --

Adding new feeling classes

To extend the emotional spectrum covered by the GEMS, we examined the utility of adding eight new feeling classes, as shown in Table 2.

-- Insert Table 2 here --

The first two classes pertain to aesthetic and epistemic emotions. We have argued elsewhere that these two types of emotions differ from utilitarian emotions in the nature and importance of the eliciting objects and events, the central appraisal criteria involved, and the nature of the response patterns in the different components (Scherer, 2004; Scherer & Coutinho, 2013). In our perspective, a major difference between utilitarian emotions on the one hand and aesthetic and epistemic emotions on the other is the fact that appraisals concerning goal relevance and coping ability (which are central criteria for utilitarian emotions) involve different criteria (such as different goals and coping mechanisms). In other words, an aesthetic or epistemic emotional experience is not triggered by concerns with the immediate relevance of an event for one's survival or well-being, nor with how well one can cope with the situation. Rather, of paramount importance is the appreciation of the intrinsic qualities of a piece of visual art or a piece of music, or the degree of discovery or insight one achieves through novel and complex stimulation in different modalities.

This view is, in essence, similar to Brattico, Bogert and Jacobsen's (2013) suggestion that aesthetic emotions elicited by music have no survival function and result from intrinsic appreciation of the music at different levels. Brattico and colleagues further specify that aesthetic emotions are elicited in specific circumstances (e.g., aesthetic modes of listening), and may follow the induction of

other, non-aesthetic emotions (including discrete emotions). In relation to this point, the question is whether aesthetic emotions are independent/different of other emotions (a separate class) or, instead, blends/by-products of other affective states (e.g., Barrett, Grim, Robins, Wildschut, Sedikides, & Janata, 2010). Trost, Ethofer, Zentner, & Grandjean (2012) suggest that the emotions described by some of the GEMS scales are aesthetic in nature (in that they lead to differentiated brain patterns), and that aesthetic emotions may consist of blends with more basic affective states. This is an area under active debate and any attempt to find a consensual definition of aesthetic emotions at this stage seems doomed to failure. Thus, some authors posit that aesthetic emotions are the only type of emotions that can be induced by music (Konecni, 2008), whereas others suggest that they are only a particular type of emotions that music can induce through an appreciation of its intrinsic qualities (Omigie, 2015; Brattico et al. 2013; Juslin, 2013; Scherer & Coutinho, 2013). Here we considered a distinctive variety of aesthetic emotions elicited by specific aesthetic judgments: beauty and harmony. Our goal is to determine to what extent the experiential qualities associated with these judgments ("feelings of beauty" and "feelings of harmony") form a coherent cluster and whether they are already captured by other GEMS dimensions or describe a different set of feelings.

Although epistemic and aesthetic emotions are typically not distinguished in the literature, we also included epistemic emotions as a distinctive class and evaluated whether this is separable from aesthetic emotions (and other subscales). In our view, epistemic emotions might be considered a separate (although related) set of emotions compared to aesthetic ones due to the fact that they derive from evaluations of auditory or visual stimuli in terms of their information content and contribution to knowledge or insight (in contrast with intrinsic qualities of form or relationship of elements associate with aesthetic judgments). To describe feelings related to

epistemic emotions we included the terms “interested”, “discovering novelty”, and “insight” in our study.

The next new feeling class is “Moved.” It comprises the term “moved,” which is part of the GEMS “Wonder” factor (loading of .75). However, in our view, “moved” is not ideally suited to represent this factor. The term “moved” is often used, in different contexts, to describe an affective experience involving some sort of emotional engagement with unclear or multiple emotional qualities (see Hanich, Wagner, Shah, Jacobsen, & Menninghaus, 2014). Furthermore, it is often considered to be an independent feeling class (Bartel, 1992; Juslin & Laukka, 2004; Konecni, 2008). The term “touched”, arguably referring to a similar feeling, was chosen as a complement for the “Moved” class.

The next two classes are “Boredom” and “Enthusiasm.” They refer to types of feelings that seem to be particularly important in the context of contemporary art music. Indeed, instances of these two classes were frequently reported during the pilot studies referred to above. The feeling class “Boredom” is represented by the terms “indifferent” and “bored.” The former term is one of the low-frequency items (reported by 4.6% of the respondents). The latter term is not reported in the GEMS-T2. The class “Enthusiasm” includes the terms “passionate,” “enthusiastic,” and “in awe.” Only the first term was reported in GEMS-T2 (with a weighted frequency of 23.4%).

Finally, we also tested the inclusion of three other feeling classes related to the so-called basic emotions: “Fear,” (apprehensive, uneasy),“Surprise,” (astonished, amazed) and “Anger” (aggressive, irritated, angry) (see Zentner et al., 2008 and Laukka, 2007). These three classes are often used in studies of musical emotions despite the fact that they are not commonly experienced in response to music. Nevertheless, given that other basic emotions are also included in the original GEMS

(joy and sadness), we decided to include these classes to determine to what extent they overlap with the current GEMS factors.

Empirical development of the new checklist

The purpose of the empirical study described below was to examine the *conceptual structure* of the feeling classes and the constituent terms in order to facilitate and justify the choice of classes and terms for the final version of the new checklist (GEMIAC). Specifically, we aimed at 1) establishing the coherence of the postulated feeling classes in terms of the constituent terms (and to choose two items in those cases in which three candidates had been nominated) and 2) examine the degree of overlap between the different feeling classes. We consider it important to examine these issues in terms of the general conceptual structure of the chosen terms in the English language (using similarity judgments by native speakers) rather than in concrete musical listening situations because the subjective experiences in different musical contexts are likely to bias the participants' responses. The principle of a fuzzy-set definition of feeling classes by two terms requires that the semantic structure of the checklist will be perceived in the same way by speakers of a particular language, independent of context.

The aim of the study described here is to obtain consensual judgments (in the form of arithmetic means as units of analysis) for the conceptual structure of the relationships between the chosen terms. In consequence, the number of participants required for the establishment of stable means is determined by the reliability of the judgments rather than by power consideration (as in the case of studies in which individuals are the units of analysis).

Method

Participants

Based on the evidence from large-scale earlier studies on semantic conceptual structures of emotion terms (Fontaine, Scherer, & Soriano, 2013, Table 6.2, p.103), we recruited a total of 18 subjects (13 female) with ages ranging from 21 to 45 years old (M = 30.2, SD = 6.19) . Participants were students (10) and collaborators (8) from the University of Geneva with English as native language. Each participant received a compensation of CHF 40 for participating in the study.

Materials

A list of 44 terms describing 17 classes of affective feelings (see Tables 1 and 2) was used in this study.

Procedure

The rating sessions were administered online. Participants were recruited via email, and those who agreed to participate received another email with links to the online study. Participants were asked to rate the dissimilarity between affect terms shown in Table 1 presented pairwise in all possible combinations (details below), in total 946 (44 × 43/2) ratings. Because of the high number of ratings, the task was distributed over two sessions (lasting maximally 50 min each). Participants were shown the full set of terms to be rated at the beginning of the first and second sessions, and the order of presentation of each pair of terms was randomized for all participants.

For each pair of terms, participants were asked to rate the dissimilarity between the emotional experiences associated with each of the two items in the imagined context of music listening, that is, imagining emotions felt while listening to music. Ratings were performed by using a computer mouse to quantify the distance between two rectangular boxes containing each of the two emotion terms as an indicator of the closeness of their conceptual relationship. Thus, when the boxes were placed close to each other, the emotional experiences while listening to music as described by each

term were rated as similar. The greater the distance between the two boxes the larger the presumed difference in meaning between the terms (dissimilarity). The scale ranged continuously from 0 (*minimum dissimilarity*; boxes together) to 1 (*maximum dissimilarity*; boxes maximally apart).

Results

The ratings provided by the participants were organized into 18 different matrices, one for each participant. All matrices are symmetrical and square (44 rows/columns), and each line and column correspond to one particular emotion term. The values of the diagonal were considered to be 0 (identity). Inter-rater reliability was computed by using Cronbach's alpha. Ratings were highly concordant across participants ($\alpha = .96$), justifying the assumption of convergence toward a stable mean with N= 18. In consequence, the responses were aggregated by calculating the arithmetic mean across all participants for each of the ratings (to be used as units of analysis) and stored in a mean dissimilarity matrix.

Our goal was to investigate the semantic space formed by the 44 feeling terms to evaluate whether the hypothesized fuzzy two-item feeling classes are consistent, i.e., the pairs of terms forming coherent clusters. To this effect, we conducted an agglomerative hierarchical clustering analysis (HCA; Kaufman & Rousseeuw, 1990) on the dissimilarity matrix containing the pairwise dissimilarity ratings between all pairs of terms by using the Unweighted Pair Group Method with Arithmetic Mean algorithm (Sokal & Michener, 1958; method implemented in R, *cluster* package, *agnes* function with parameter *method* set to *average*; R Core Team, 2013).

HCA organizes elements into hierarchically organized cluster levels which can be presented as tree diagrams (a.k.a. dendrograms) in which branches represent the clusters at different levels. The process of selecting specific branches or clusters as being of interest is typically referred to as branch or tree pruning. The most

commonly used method for tree pruning is the so-called fixed height branch cut. This method implies that the researcher defines a fixed height on the dendrogram, and each adjoining branch of elements below the specified height is identified as an independent cluster. Unfortunately, the solutions obtained with this method depend very much on the cut height chosen, and small changes in this value can lead to very different solutions. Further, while distinct clusters can be easily detected by visual inspection, the computational clusters determined by static cut methods often do not identify clusters correspondingly.

To address this limitation, Langfelder, Zhang and Horvath (2007) have developed new dynamic branch cutting methods based on the analysis of the shape of the branches in a dendrogram. In particular, the Dynamic Hybrid (DH) algorithm builds clusters from bottom up, using information from the dendrogram as well as the dissimilarity among elements to improve the detection of outliers in each cluster. This algorithm is implemented in two steps. In step 1, branches that satisfy specific criteria for forming clusters are identified. These criteria are: having a minimum number of member objects; excluding elements that are too far from a cluster (even if they belong to the same branch of the dendrogram); clusters being separated from surrounding clusters by a gap; and the lowest-merged objects in the cluster being tightly connected. In step 2, elements that were not assigned to any cluster in step 1 are tested for sufficient proximity to previously identified clusters. If the closest cluster is close enough the previously unassigned element is assigned to that cluster, otherwise the element is not assigned to any cluster (therefore not fitting the structure of the hierarchical solution found). In this step only dissimilarity information is used, essentially applying a modified version of k-medoid partitioning (a.k.a Partitioning Around Medoids, PAM; Kaufman & Rousseeuw, 1990) - the distance between points belonging to a specific cluster and a point designated as the center of that cluster is

minimized. We chose to employ this DH method for the identification of semantic clusters in our data.

The results of the HCA and DH algorithms for cluster identification are shown in the dendrogram reproduced in Figure 1 (agglomeration coefficient = .79). The bottom array of colors identifies the originally proposed groupings, whereas the top array labels the clusters identified by the DH algorithms with the same color. The following criteria were used to determine a preselection of the most relevant clusters and the respective pair of terms on the basis of our data:

1) Each cluster must contain two and only two terms
2) Terms not assigned to any cluster (i.e., standing alone) are removed
3) When more than two terms are assigned to a single cluster, the two closest terms (i.e., the latest to be separated in the tree diagram) with the lowest break points are kept and the remaining are removed

-- Insert *Figure 1* here --

Preselection of the list of feeling classes and terms to be included in the checklist

For the sake of economy, our detailed considerations concerning the application of these criteria and the final choices concerning the pairs of terms for the different fuzzy feeling classes are provided in Section B of the Supplementary Material to this article. In particular, we fine-tuned the terms to represent the original 9 GEMS factors. For this purpose we had to eliminate some terms from further consideration as additional terms had been added as potential alternatives and to decide on the final list of pairs to represent the 9 GEMS factors. The final choice of items is listed in Table 3 (correspondence between GEMS factors and new clusters indicated in the last column). We refrain from providing overarching labels for these classes as experience

has shown that such headings are often not compatible with the notion of a fuzzy class as they introduce additional and often discrepant meaning.

-- Insert Table *3* here --

In addition, we had to examine the feeling classes we wanted to add as an extension to the coverage of the GEMS in terms of the coherence of the terms proposed as designating the targeted feeling classes and their relationships to other feeling classes, including the classes representing the GEMS factors. Again, further details concerning this selection procedure are reported in Section B of the Supplementary Material.

In addition to the subscales initially considered, two strong new clusters emerged from our analysis which we considered as potential subscales to be added given their pertinence. The first cluster comprises W2 ("enchanted") and E3 ("in awe"), terms initially proposed for the "Wonder" and "Enthusiasm" subscales (respectively). Together with our previous findings in relation to the "Wonder" and "Moved" classes (see Section B of the supplementary materials), this new cluster indicates that the original GEMS factor "Wonder" was rather complex and contained multiple feelings which differ significantly in terms of experiential qualities. The second emergent cluster includes the terms P2 ("energetic") and J1 ("lively"). Albeit similar to the "Joyful activation" cluster, these two terms specifically describe the energizing effect of music, which are independent of the action tendencies associated with "joy" and "wanting to dance."[1] A recapitulation of the new classes added to the

[1] It should be noted that during the development of the GEMS there was evidence for the existence of two separate clusters – one related to activation and another to joy. This seems to be evident also in our work. Whereas the GEMS's authors have opted to join both factors (see p. 505), our results suggest their separation.

nine redefined GEMS classes. represented by the corresponding terms, is shown in Table *3*.

Semantic structure of the final list of terms

Having removed the terms that were not selected for inclusion in the checklist (as described in the above analysis and the Supplementary Material) from the initial list of candidates, we repeated the HCA and clustering procedure using the dissimilarity matrix where the rows and columns pertaining to the discarded terms were removed (the final matrix comprises the dissimilarities amongst 24 terms). The resulting dendrogram is shown in Figure 2.

-- Insert *Figure 2* here --

As can be seen, the clustering solution shown in Figure 1 confirms the existence of a stable structure containing 12 independent clusters comprising two feeling terms each. Furthermore, our changes in the list of terms also lead to a more robust clustering structure, as noted by the increase in the agglomeration coefficient (.81 in the second solution vs .79 in the initial one).

Discussion

Generally, our results indicate that all original GEMS factors form consistent clusters of emotional meaning. This was expected given the comprehensive data collection and analyses behind its development. Indeed, albeit with some changes in the specific terms used for some of the scales, the GEMS factors of Transcendence, Tenderness, Nostalgia, Peacefulness, Power, Joyful activation, Tension and Sadness, emerged from the hierarchical clustering analysis. Arguably, the minor changes in

specific terms used for each subscale (called factors in the original GEMS), improved the semantic independence between scales. Some of the observed patterns deserve further attention.

The "Wonder" factor in GEMS (reported as ambiguous by some participants in empirical studies; cf. introduction section) expanded to three more specific and independent subscales in this study: "Wonder" ("filled with wonder" and "amazed"), "Being moved" ("moved" and "touched") and "Awe" ("enchanted" and "in awe"). In particular, the "wonder" subscale is now more specific compared to the original GEMS factor, and it clearly links "wonder" to the concept of surprise (as noticed earlier the proposed terms for the surprise subscale are describe similar emotional experiences to those described as wonder; this as also been noted by others, e.g., Frijda, 1986, and is a common element in the definition of wonder). "Moved", as expected and in accordance to other literature (Bartel, 1992; Juslin & Laukka, 2004), forms an independent subscale. Finally, a new subscale emerged that combined two terms initially hypothesized for the "wonder" and "enthusiasm" subscales – "enchanted" and "in awe", respectively. Compared to the "wonder" subscale they differ (at least) in the underlying element of surprise associated with the respective terms. This subtle (albeit important) distinction is compatible with Konecni's (2005) view, according to which awe is a "prototypical subjective reaction to a sublime stimulus" and the "ultimate, aesthetic response", which can be accompanied by feelings of wonder and being moved. Further theoretical and empirical discussions of the concept of awe in general and in relation to music in particular can be found in Konecni (2005), Keltner & Haidt (2003) and Shiota, Keltner, & Mossman (2007). In this paper we will refer to this new scale as "Enchanted, in awe".

In addition to "Moved, touched" (initially proposed) and "Enchanted, in awe" (emerging from the analysis) subscales, we found evidence for the inclusion of three

new subscales, which we will refer to as “Energetic, lively”, “Indifferent, bored” and “Agitated, angry”. “Energetic, lively” is related to the “Joyful activation” GEMS factor, but emerged as a specific scale in our study. The terms “energetic” and “lively” are often reported in mood regulation studies (Thayer, Newman, & McClain, 1994). As mentioned before, “Energetic, lively” and “Joyful, wanting to dance” are similar in terms of arousal, but can be differentiated with respect to the actions tendencies associated with joy and dancing. On the negative side of the emotional spectrum, “indifferent, bored” and “agitated, aggressive” emerged as semantically relevant clusters. The first clearly describes feelings of disengagement with the music, whereas the second described emotional responses related to action tendencies related to irritation or anger.

Our analysis failed to find supporting evidence for four other proposed subscales -- “Aesthetic feelings”, “Epistemic feelings”, “Surprise” and “Fear”. As described before, the terms hypothesized for these clusters were either overlapping with other established clusters in the semantic space, or the various terms were strongly overlapping with other terms in the analysis. In relation to “Aesthetic feelings”, “beauty” and “harmony”, were clustered with the “Moved, touched” and “Relaxed, peaceful” subscales. The “Epistemic feelings” terms “discovering novelty” and “interested”, were clustered with the “Filled with wonder, amazed” and “Enchanted, in awe” subscales, respectively. This suggests, in line with Trost et al. (2012), that the latter two subscales are related to aesthetic emotional responses. However, it is interesting to observe that none of the terms hypothesized for the aesthetic and epistemic feeling subscales were clustered together, which suggests that our earlier categorization may be meaningful in some contexts. In relation to “Surprise”, our analysis showed a strong overlap with the “Filled with wonder, amazed” subscale (which is now formed by one term initially proposed for a

"Surprise" subscale). Finally, we found no evidence for a "Fear" cluster using the terms proposed. "Uneasy" overlapped with "Tension" subscale (and we consequently chose "Uneay, tense" for that subscale), and "apprehension" was eliminated from the analysis due to the fact that did not fit the clustering solution in the context of the criteria set for the inclusion.

We believe that the results reported here, showing a well-formed hierarchical structure of the selected terms, justifies the use of two-item fuzzy affect classes in the checklist. In consequence, the 14 fuzzy-set classes shown in Table 3 constitute Version 1.0 of the GEMIAC. How can this checklist be validated over and above the demonstration of stringent semantic structure properties in this article? Given that we use two item fuzzy-set subscales, the normal computation of scale reliability (e.g., using Cronbach's alpha) is not applicable. As to face validity, we have made frequent references to other literature showing a high degree of parallelism with respect to dimensions and terms. Construct validation is difficult to establish as there are few other instruments that are directly comparable (as the GEMS was used as the basis for the GEMIAC development it cannot be used for construct validation). The establishment of predictive validity is also difficult to envisage given that there seem to be no established criteria in terms of objective outcome measures or behavioral indices. Validation in the sense of predictive validity is particularly difficult as there are extraordinary individual differences with respect to affective reactions, even to the same piece of music, which makes it impossible to develop an empirical indicator for an underlying latent variable. It would seem that the value and utility of the checklist will need to be established through extensive use in empirical research projects, especially in field settings (in all types of concert venues, opera houses, music festivals, churches, clubs, etc.). Among the criteria of utility could be the ability of the GEMIAC scales to differentiate the emotional reactions to different genres of music,

of different themes and movements, and, in particular, of different types of interpretation. In general, many aspects of the route model of emotional induction (Scherer & Zentner, 2001; Scherer & Coutinho, 2013) as well as the BRECVEM (Juslin, Liljeström, Västfjäll, & Lundqvist, 2010) and its offshoots (Juslin, 2013), could be operationalized in terms of hypotheses bearing on the GEMIAC feeling classes.

In general, given that the GEMIAC is closely modeled on the GEMS, it would seem that the extensive usage of the latter, testified in many publications, would suggest a high degree of utility and validity also for the former. There are only two aspects that are different with respect to the 9-scale brief of the GEMS – 1) the use of two-term fuzzy set descriptors for the emotion classes as compared to one overarching label, and 2) five additional emotion classes. In the process of the GEMIAC development we have taken care to examine these issue with respect to the acceptance of the fuzzy-set rating categories by listeners and the ability of the additional scales to differentiate the emotional reactions to different genres of music, of different themes and movements within one musical work, and, in particular, of different types of interpretation. For this purpose, preliminary versions of the GEMIAC have been used extensively in our group for specific studies, e.g., 1) to examine the emotional reactions to public opera performances (Trznadel, Fantini, & Scherer, submitted; demonstrating differentiation between operas and acts and scenes within a specific opera; 2) studying the affective reactions to Schubert Lieder during a recital in a church setting (Coutinho & Scherer, submitted; results showing clear differences between Lieder with different affective tones); and 3) measuring moment-to-moment affective reactions of the audience during the performance of three pieces – Haydn, Ligeti, Schubert -- by a professional string quartet in a concert hall (Beermann et al., in prep.; clearly demonstrating differences in affective reactions

between the classical and the contemporary piece). All three studies showed strong acceptance of the fuzzy-set rating scales by the participants.

One further issue of interest is the dimensional structure of the GEMIAC. The examination of the semantic structure of the GEMIAC revealed some interesting findings in this respect. Negatively valenced feelings of emotions (melancholic and sad, nostalgic and sentimental, indifferent and bored, tense and uneasy, agitated and aggressive) shows a clear semantic separation from positive terms at the highest hierarchical level. The next sharp division of the negative valenced branch, separates terms in terms of arousal. Terms related to anger and tension (high arousal) appear in a single sub-branch, whereas terms characterized by low arousal (melancholic, sad, nostalgic, sentimental, indifferent and bored) are clustered in another sub-branch. The positive valenced top branch is first subdivided into feelings characterized by low arousal (full of tenderness and warmhearted, nostalgic and sentimental) and high arousal (remaining terms) feelings. Within those terms associated with positive valence and high arousal there is another interesting separation, which represents a division between emotions primarily concerned with the self and inner feelings (inspired and enthusiastic, energetic and lively, joyful and wanting to dance, powerful and strong) and those that are more related to the music itself (filled with wonder and amazed, moved and touched, enchanted and in awe). This last group is reminiscent of the concept of self-transcendent positive emotions proposed by Schindler et al. (2013). More generally, this semantic structure also reflects the differentiation between utilitarian and aesthetic emotions (as defined in the introduction to this article). The former are close to everyday emotions that serve to adapt the individual to important events, whereas the latter are related to the appreciation of the intrinsic qualities of a piece of music or the degree of discovery or insight one achieves

through novel and complex stimulation in different modalities (see Scherer & Coutinho, 2013, for a discussion).

Finally, it is relevant to mention that the question posed to the participants was how similar (or dissimilar) the feelings of emotion described by each pair of terms were in the context of music listening. Therefore, the groupings resultant for the clustering analysis are based on the feeling qualities that the terms/expressions describe rather than the definition per se. This is in our view a strength of our approach, and not a limitation – in a particular scale, those elements that define the same latent construct should be as closely related as possible in the context of the construct that they are intended to measure. In this context it is worth mentioning that in preliminary stages of development we conducted a similar empirical study with a different aim (asking about the similarity between the feelings of emotion described by various pairs of terms, but in the context of everyday life experiences that not music listening) and using the same vocabulary. We found that the organization of the terms was rather different. This supports the idea that the terms are being group based on the experiential qualities of the feelings they describe in the context of music listening, rather than their definition.

Conclusions and Outlook

The development of appropriate methods for measuring and classifying emotional states is of paramount importance in music research, for they are used to generate concrete hypotheses related to the emotional power of music. Nevertheless, to date, only one instrument has been systematically developed on an inductive basis to measure musically induced emotions: the GEMS. In this article, we propose a complementary instrument, a checklist that aims to extend the affective spectrum captured by the GEMS by providing robust differentiation between the emotional

feelings considered and facilitating the rapid assessment of musically induced affect in a wider range of experimental contexts, especially in field studies.

In particular, in order to obtain audience collaboration in music events, the list must be very short and easily fit on one page with instructions. This format precludes the use of traditional scales consisting of theoretically postulated groups of items that can be analyzed by reliability estimates and confirmatory factor analysis. This process was applied for the development of the GEMS, but experience has shown that such a long scale can be used in laboratory settings but becomes prohibitive in field studies. Therefore, the development of GEMIAC involved the empirical analysis of the hierarchical semantic relationships between various affect terms and the identification of those pairs of terms that formed coherent feeling classes (i.e., the semantic distance between two terms in the same feeling class, which is smaller than the distance to any other feeling term in the checklist) depicting affective states commonly experienced whilst listening to music. The results of our empirical study with a set of 44 terms led to a final checklist comprising 14 fuzzy categories, including the well-established nine dimensions of the GEMS and five experimentally confirmed additional feeling classes related to being moved (“moved”, “touched”), enthusiasm (“inspired”, “enthusiastic”), energy (“energetic”, “lively”), disengagement (“bored”, “indifferent”), and anger (“agitated”, “aggressive”). Each (fuzzy) category entails two affective terms that are semantically closely related.

We invite researchers in the area of music and emotion to examine the utility of the GEMIAC, which is freely available for non-commercial research, in their work. Two forms of the final instrument, for both intensity and frequency measurements, are reproduced in Supplemental Material Sections C and D. Intensity measurement is recommended in cases in which listeners hear a short piece of music with a relatively homogeneous emotional tonality and rate the feelings induced immediately after the

experience, whereas measurement of the frequency with which a certain feeling class has been experienced is more appropriate when listeners assess their affective reaction after a long work of music, such as an opera, with many different emotional qualities occurring over time. We consider the GEMIAC as an "open" assessment instrument in that we invite researchers to add two-term fuzzy affect classes to the 14 classes currently proposed to allow an adaptation of the instrument to specific research aims and, in particular, musical genres that have been neglected up to now. However, we suggest to carefully examine, whenever possible empirically, that the two terms describing a fuzzy class are indeed semantically highly related without being direct synonyms. Unless this is the case, listener-raters may be confused, endangering the reliability and validity of the data. Furthermore, we request that researchers always keep the 14 standard GEMIAC classes and place these at the top of the rating sheets, to ensure that conditions are comparable across different studies allowing direct comparison of the data.

We hope that this new checklist for music-induced affect, the GEMIAC, will provide a useful means for rapidly assessing different qualities of affective states induced by music of different genres for researchers from many different disciplines in different listening contexts. Given the enormous variability of affect and emotion lists used in different research projects, convergence on a minimal list of relevant affect classes and a standard rating format would greatly increase our cumulative knowledge in this dynamic and expanding research field.

References

Aljanaki, A., Wiering, F., & Veltkamp, R. C. (2016). Studying emotion induced by music through a crowdsourcing game. *Information Processing & Management*, *52*(1), 115-128.

Barrett FS, Grimm KJ, Robins RW, Wildschut T, Sedikides C, Janata P. 2010. Music-evoked nostalgia: affect, memory, and personality. Emotion. 10:390--403.

Beermann, U., Trznadel, S. & Scherer, K.R. (in prep.) Moment-to-moment tracking of affective reactions to chamber music in a concert hall setting.

Brattico E, Bogert B and Jacobsen T (2013) Toward a neural chronometry for the aesthetic experience of music. *Front. Psychol.* **4**:206. doi: 10.3389/fpsyg.2013.00206

Coutinho. E. & Scherer, K. R. (under revision). The effect of context and audio-visual modality on emotions elicited by a musical performance.

Damasio, A., & Carvalho, G. B. (2013). The nature of feelings: evolutionary and neurobiological origins. *Nature Reviews Neuroscience*, *14*(2), 143-152.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, *6*(3–4), 169–200.

Fontaine, J.R.J., Scherer, K.R., & Soriano, C. (Eds.). (2013). *Components of emotional meaning: A sourcebook*. Oxford: Oxford University Press.

Frijda, N. (1986). The emotions. Cambridge, UK: Cambridge University Press.

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika, 53*, 325–328.

Hair, J. F., Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis* (5th ed.). Upper Saddle River, NJ: Prentice-Hall.

Hanich, J., Wagner, V., Shah, M., Jacobsen, T., & Menninghaus, W. (2014). Why we like to watch sad films: The pleasure of being moved in aesthetic

experiences. *Psychology of Aesthetics, Creativity, and the Arts*, 8(2), 130–143. doi:10.1037/a0035690

Hevner, K. (1936). Experimental studies of the elements of expression in music. *American Journal of Psychology, 48*, 248–268.

Juslin, P. N. (2013). From everyday emotions to aesthetic emotions: Towards a unified theory of musical emotions. *Physics of Life Reviews, 10*, 235–266. doi:10.1016/j.plrev.2013.05.008

Juslin, P. N. (2013). From everyday emotions to aesthetic emotions: towards a unified theory of musical emotions. *Physics of Life Reviews*, *10*(3), 235-266.

Juslin, Liljeström, Västfjäll, & Lundqvist. (2010). How does music evoke emotions? Exploring the underlying mechanisms. In P.N. Juslin & J. Sloboda (Eds.), *Handbook of Music and Emotion: Theory, Research, and Applications* (pp. 605-642). Oxford: Oxford University Press.

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data: An introduction to cluster analysis* (1st ed.). New York, NY: John Wiley.

Keltner, D., & Haidt, J. (2003). Approaching awe, a moral, spiritual, and aesthetic emotion. *Cognition & Emotion*, *17*(2), 297-314.

Kitayama, S., Markus, H. R., & Kurokawa, M. (2000). Culture, emotion, and well-being: Good feelings in Japan and the United States. *Cognition & Emotion*, *14*, 93–124.

Kivy, P. (1980). *The corded shell: Reflections on musical expression*. Princeton, NJ: Princeton University Press.

Konecni, V. J. (2008). Does music induce emotion? A theoretical and methodological analysis. *Psychol. Aesthet. Creat. Arts* 2, 115–129.

Kruskal, J. B. 1964a. Multidimensional scaling by optimizing goodness-of-fit to a nonmetric hypothesis. *Psychometrika, 29,* 1–28.

Kruskal, J. B. 1964b. Nonmetric multidimensional scaling: A numerical method. *Psychometrika, 29,* 115–129.

Laukka, P. (2007). Uses of music and psychological well-being among the elderly. Journal of Happiness Studies, 8, 215–241

Langfelder, P., Zhang, B., & Horvath, S. (2007) Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics,* 24(5), 719-720.

McCraty, R., Barrios-Choplin, B., Atkinson, M., & Tomasino, D. (1998). The effects of different types of music on mood, tension, and mental clarity. *Alternative therapies in health and medicine*, *4*(1), 75-84.

Omigie, D. (2015). Basic, specific, mechanistic? Conceptualizing musical emotions in the brain. *J. Comp. Neurol.* doi: 10.1002/cne.23854 [Epub ahead of print].

R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/.

Scherer, K. R. (2004). Which emotions can be induced by music? What are the underlying mechanisms? And how can we measure them? *Journal of New Music Research, 33*, 239–251.

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information, 44,* 693–727.

Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and Emotion, 23(7),* 1307-1351.

Scherer, K., & Coutinho, E. (2013). How music creates emotion: A multifactorial process approach. In T. Cochrane, B. Fantini, & K. R. Scherer (Eds.), *The emotional power of music*. Oxford, UK: Oxford University Press.

Scherer, K. R., & Zentner, K. R. (2001). Emotional effects of music: Production rules. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 361–392). Oxford, UK: Oxford University Press.

Schindler, I., Zink, V., Windrich, J., & Menninghaus, W. (2013). Admiration and adoration: Their different ways of showing and shaping who we are. *Cognition & Emotion*, *27*, 85–118. doi:10.1080/02699931.2012.698253

Shiota, M. N., Keltner, D., & Mossman, A. (2007). The nature of awe: Elicitors, appraisals, and effects on self-concept. *Cognition and Emotion*, *21*, 944–963.

Sloboda, J.A., & O'Neill, S.A. (2001). Emotions in everyday listening to music. In: P.N. Juslin, & J.A. Sloboda (Eds.), *Music and Emotion: Theory and Research*. New York: Oxford University Press, 415–430.

Sokal, R., & Michener, C. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin*, 38, 1409–1438.

Sturrock, K., & Rocha, J. (2000). A multidimensional scaling stress evaluation table. *Field Methods*, *12*, 49–60.

Thayer, R. E., Newman, J. R., & McClain, T. M. (1994). Self-regulation of mood: Strategies for changing a bad mood, raising energy, and reducing tension. *Journal of Personality and Social Psychology*, *67*, 910–925. doi: 10.1037/0022-3514.67.5.910.

Torres-Eliard, K., Labbe, C., & Grandjean, D. (2011). Towards a dynamic approach to the study of emotions expressed by music. In *Proceedings 4th international ICST conference on intelligent technologies for interactive entertainment* (pp. 252–259).

Trost, W., Ethofer, T., Zentner, M., & Vuilleumier, P. (2012). Mapping aesthetic musical emotions in the brain. *Cerebral Cortex*, *22*(12), 2769-2783.

Trznadel, S., Fantini, B., & Scherer, K. R. (submitted). Emotional experiences during opera performances: A field study of the nature and determinants of affective feelings reported by spectators.

Vuoskoski, J. K., & Eerola, T. (2010). Domain-specific or not? The applicability of different emotion models in the assessment of music-induced emotions. In *Proceedings of the 10th international conference on music perception and cognition* (pp. 196-199).

Zentner, M., & Eerola, T. (2010). Self-report based measures and models. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion* (pp. 187–221). Oxford, UK: Oxford University Press.

Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion*, *8*, 494–521.

Table 1

List of Feeling Terms Pertaining to the Nine GEMS Factors (or Feeling Classes)

Feeling class	Term	Alias	In GEMS-A1?	In GEMS-T2?	Observations
Wonder	Filled with wonder	W_1	Yes	Yes	
	Enchanted	W_2	No	Yes	
Transcendence	Feelings of transcendence	Tr_1	Yes	Yes	
	Inspired	Tr_2	Yes	Yes	
Tenderness	Full of tenderness	Te_1	Yes	Yes	
	Warmhearted	Te_2	**No**	**No**	Synonym for feeling of affection
Nostalgia	Melancholic	N_1	Yes	Yes	
	Nostalgic	N_2	Yes	Yes	
	Dreamy	N_3	Yes	Yes	
	Sentimental	N_4	Yes	Yes	
Peacefulness	Relaxed	Pe_1	Yes	Yes	
	Soothed	Pe_2	Yes	Yes	
	Peaceful	Pe_3	Yes	Yes	Factor name
Power	Powerful	P_1	Yes	Yes	Factor name
	Energetic	P_2	Yes	Yes	
	Strong	P_3	Yes	Yes	
Joyful activation	Lively	J_1	**No**	**No**	Synonym for “animated”
	Joyful	J_2	Yes	Yes	
	Wanting to dance	J_3	Yes	Yes	Reformulated from “feeling like dancing”
Tension	Tense	T_1	Yes	Yes	
	Nervous	T_2	**No**	Yes	
	Aroused	T_3	Yes	Yes	
	Agitated	T_4	Yes	Yes	
Sadness	Sad	S_1	Yes	Yes	
	Gloomy	S_2	**No**	**No**	

Note. GEMS = Geneva Emotional Music Scale; GEMS-A1 = GEMS reported by Zentner and colleagues (2008, Table A1, p. 519); GEMS-T2 = GEMS reported by Zentner et al. (2008, Table 2, p. 504).

Table 2

List of Newly Proposed Feeling Terms Pertaining to Eight New Feeling Classes

Feeling class	Term	Alias	In GEMS-A1?	In GEMS-T2?	Observations
Aesthetic emotions	Feelings of harmony	AE_1	No	No	
	Feelings of beauty	AE_2	No	No	
Epistemic emotions	Interested	K_1	No	No	
	Discovering novelty	K_2	No	No	
	Insight	K_3	No	No	
Moved	Moved	M_1	Yes	Yes	
	Touched	M_2	No	Yes	
Boredom	Indifferent	B_1	No	Yes	
	Bored	B_2	No	No	
Enthusiasm	Passionate	E_1	No	**Yes**	
	Enthusiastic	E_2	No	No	
	In awe	E_3	No	No	
Fear	Apprehensive	F_1	No	No	
	Uneasy	F_2	No	No	Meta-factor label
Surprise	Astonished	Su_1	No	No	
	Amazed	Su_2	No	No	In Table 1
Anger	Aggressive	A_1	No	Yes	
	Irritated	A_2	No	Yes	
	Angry	A_3	No	Yes	

Note. GEMS-A1 = Geneva Emotional Music Scale (GEMS) reported by Zentner and colleagues (2008, Table A1, p. 519); GEMS-T2 = GEMS reported by Zentner et al. (2008, Table 2, p. 504).

Table 3

Final List of Fuzzy-set Clusters and Feeling Terms for GEMIAC

Cluster	Term	Observations
1	Filled with wonder, amazed	GEMS factor Wonder
2	Moved, touched	New subscale
3	Enchanted, in awe	GEMS factor Transcendence
4	Inspired, enthusiastic	New subscale
5	Energetic, lively	New subscale
6	Joyful, wanting to dance	GEMS factor Joyful activation
7	Powerful, strong	GEMS factor Power
8	Full of tenderness, warmhearted	GEMS factor Tenderness
9	Relaxed, peaceful	GEMS factor Peacefulness
10	Melancholic, sad	GEMS factor Sadness
11	Nostalgic, sentimental	GEMS factor Nostalgia
12	Indifferent, bored	New subscale
13	Tense, uneasy	GEMS factor Tension
14	Agitated, aggressive	New subscale

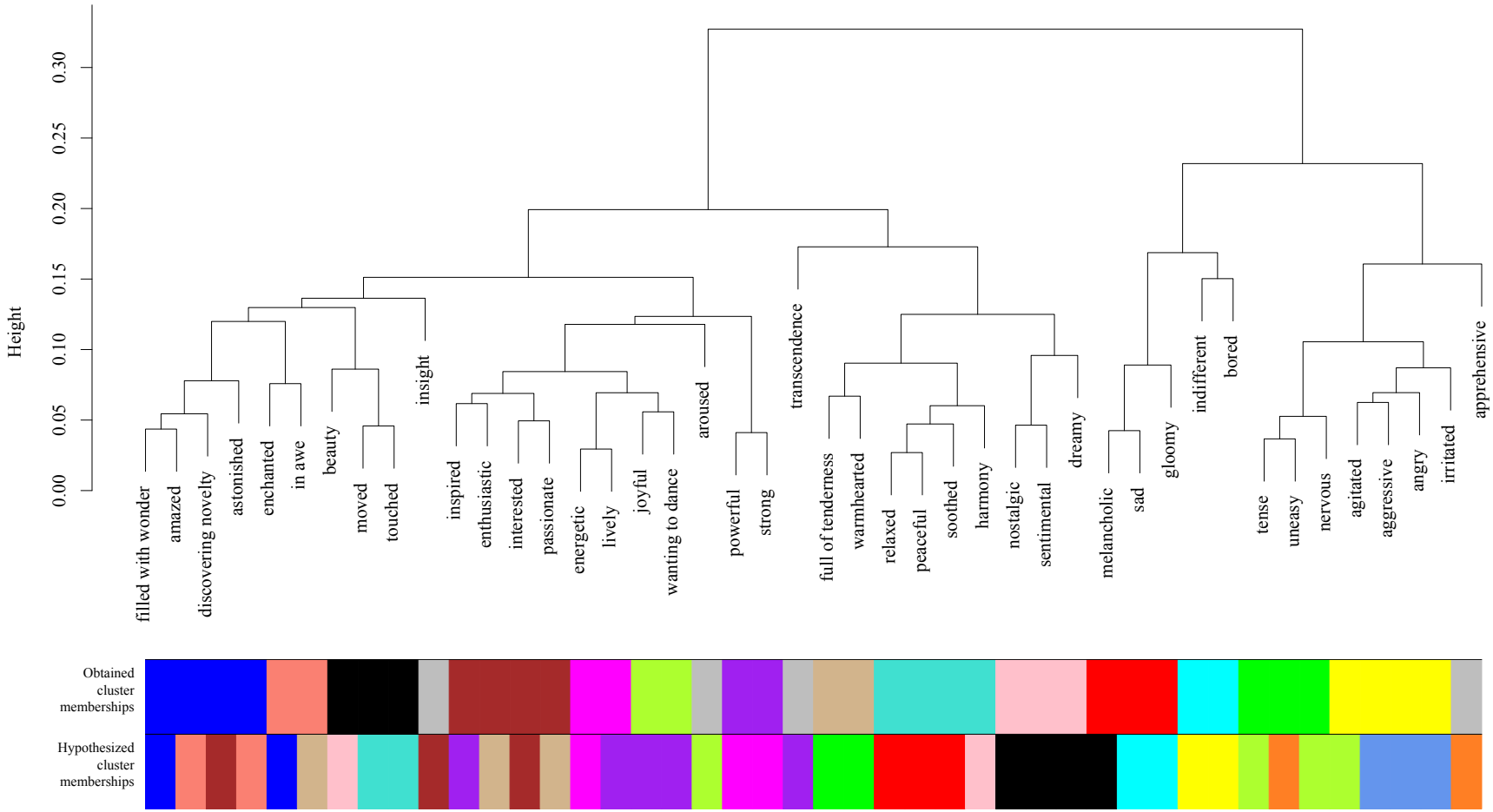

Figure 1. Semantic map of feeling terms used in our study represented in the form of a dendrogram derived from agglomerative hierarchical clustering analysis (aHCA; agglomerative coefficient = .793). The color arrays indicate the hypothesized cluster membership (bottom) and the ones obtained empirically (top).

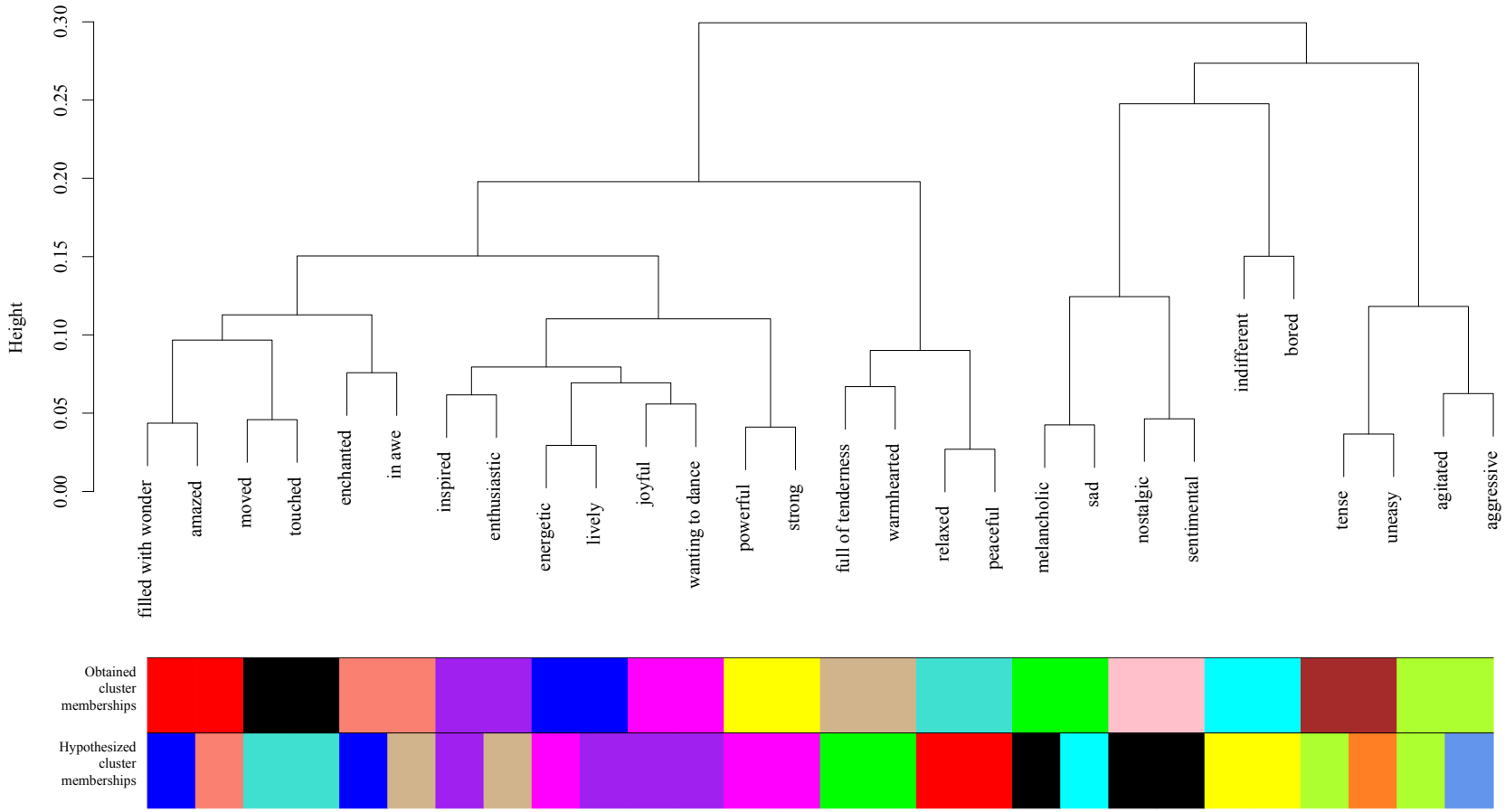

Figure 2. Semantic map of final list of feeling terms represented in the form of a dendrogram derived from agglomerative hierarchical clustering analysis (aHCA; agglomerative coefficient = .812). The color arrays indicate the hypothesized cluster membership (bottom) and the ones obtained empirically (top).

Supplemental Materials

Section A

List of terms based on the 9 factors of the GEMS scales

Wonder. We selected the terms "filled with wonder" and "enchanted" to define the "Wonder" class. The first term is part of the GEMS "Wonder" factor (factor loading of .95; see GEMS-A1). The latter term is not part of the feeling terms indicated in the GEMS "Wonder" factor, but it was one of the most frequent feelings reported by participants in the original GEMS studies (as shown in GEMS-T2), and it is commonly used as a synonym for "wonder."

Transcendence. The expression "feeling of transcendence" and the term "inspired" were chosen to represent this feeling class. Both are part of the GEMS' "Transcendence" factor (loadings of .92 and 1.00, respectively).

Tenderness. This class includes one term from the GEMS "Tenderness" factor ("tenderness", rephrased as "full of tenderness") and a synonym for "affectionate" that belongs to the same factor ("warmhearted"), both with factor loadings of .97. The term "warmhearted" was preferred to "affectionate" because it does not seem to require the assumption of a social relationship.

Nostalgia. To describe this feeling class, we selected the four terms from the GEMS' "Nostalgia" factor: "melancholic," "nostalgic," "dreamy," and "sentimental" (factor loadings of 1.00, .77, .65, and .54, respectively).

Peacefulness. This class is described by using three terms: "relaxed" (factor loading of .96), "soothed" (factor loading of .90), and "peaceful." The latter term was not part of the final GEMS "Peacefulness" factor, but it was originally used as the overarching factor label.

Power. The "Power" class comprises the terms "powerful," "energetic," and "strong." The last two terms were part of the GEMS "Power" factor (loadings of 1.00 and .70, respectively), and the first was used as the overarching factor label.

Joyful activation. We selected the terms/expressions "lively," "joyful," and "wanting to dance" to describe this class. The term "lively" is a synonym for the term "animated" used in the original GEMS "Joyful activation" factor (loading of .95). The term "joyful" is part of the same GEMS factor with a loading of .99. Finally, "wanting to dance" is a reformulation of the GEMS' term "feel like dancing" (factor loading of .72), which is also part of the "Joyful activation" factor.

Tension. Three of the original GEMS terms pertaining to the "Tension" factor were selected to create this new class: "tense," "nervous," and "agitated" (factor loadings of .63, .85, and 1.00, respectively). Additionally, we included the term "aroused" to represent the activation component.

Sadness. We chose the terms "sad" and "gloomy" to describe the "Sadness" feeling class. The term "sad" is part of the GEMS "Sadness" factor (loading of 1.00), whereas "gloomy" is a new term that was chosen to describe a state of deep sadness closer to depression. This term seems preferable to "sorrowful" (factor loading of .82 in the original GEMS "Sadness" factor), which seems to have a connotation of personal loss.

Section B

Preselection of feeling classes and terms on the basis of the cluster analysis results reported in the article

In the following paragraphs we describe our findings for each subscale tested with the aim of identifying pairs of terms forming stable semantic clusters (for the sake of simplicity, we refer to each term by using its alias as shown in Tables 1 and 2 of the article) as well as our considerations justifying the final selection.

Semantic structure of the subscales derived from the original 9 GEMS factors

Wonder. The terms W1 and W2 do not form a consistent cluster. Instead, Su2 is more similar to W1, whereas E3 is more similar to W2. Considering that W1 was in the original GEMS (W2 was not), this term was kept to describe the "Wonder" subscale. Su2 was selected for being the closest term to W1.

Transcendence. The items hypothesized for this subscale are clearly very different. Tr1 was not assigned to any cluster which indicates that it does not fit the clustering solution found by DH). According to our criteria this item was removed. Tr2 was assigned to a cluster comprising E1, E2 and KF2, indicating that it overlaps with two other proposed feeling classes. Considering that Tr2 and E2 are the less dissimilar terms in this group, they were selected to form the Transcendence cluster.

Tenderness. The terms Te1 and Te2 form a stable cluster, and are therefore chosen to form the "Tenderness" subscale.

Nostalgia. From the four terms hypothesized for this subscale, N2, N3 and N4 form a stable cluster (N1 was clustered with S1). Given the proximity between N2 and N4, these two terms were selected to form the Nostalgia subscale.

Peacefulness. The three tested terms for this subscale were identified as a single cluster, together with AE1. According to our proximity criterion, Pe1 and Pe3 were selected to form this sub-scale, and Pe2 and AE1 were removed.

Power. The terms P1 and P3 are clustered together and form a clear subscale. The term P2 is substantially different and forms another cluster with term J1 (to be discussed later).

Joyful activation. The terms hypothesized for this subscale that form a consistent cluster are J2 and J3 and were thus selected to define the "Joyful activation" subscale. J1 is also similar to these terms, but clustered independently with P2, a term initially hypothesized for the "Power" subscale.

Tension. From the four terms tested, only T1 and T2 are part of the same cluster. This cluster includes also the term F2, which was hypothesized for the "Fear" subscale. Given the lower dissimilarity between the terms T1 and F2, and the fact that F2 is a synonym of T2, these two terms were to define the "Tension" subscale. T3 was removed since it was not part of the clustering solution. T4 is part of the "Anger" cluster (discussed below).

Sadness. The terms S1 and S2, hypothesized for the "Sadness" subscale, together with N1 (hypothesized for the "Nostalgia" subscale) from an independent cluster; nevertheless, N1 is the most similar to S1, thus forming the most cohesive cluster.

Semantic structure of the new subscales proposed as an extension of the GEMS coverage

Aesthetic emotions. The terms conjectured for this subscale are inconsistent. AE1 was identified as part of the "Peacefulness" cluster, whereas AE2 is part of the newly proposed "Moved" cluster. Both terms (and the subscale) were therefore removed from our list of terms.

Epistemic emotions. The three terms chosen for this subscale appear in very different places in the dendrogram and therefore do not form a consistent subscale.

KF1 was assigned to the "Transcendence" cluster, KF2 to the "Wonder" cluster, and KF3 was not assigned to any cluster. All three terms were removed from the list.

Moved. As expected, the term M1 did not cluster with W2 as in the original GEMS. It did form an independent cluster with the newly proposed term M2. Both terms were kept, as well as the respective subscale "Moved".

Boredom. The terms B1 and B2 form an independent cluster, and fulfill the requirement of an independent subscale.

Fear. The terms F1 and F2 do not cluster together. As mentioned earlier, F2 is closer to the "Tension" subscale (and was added to it), whereas F1 was not assigned to any cluster. F1 was removed from the list of terms, as well as an independent "Fear" subscale.

Surprise. The terms Su1 and Su2 were clustered together, but as part of the "Wonder" subscale. Su2 was kept and added as the second term of this scale, and Su1 was removed. These results indicate the "Wonder" subscale is strongly related to surprise. No support was found for an additional subscale.

Enthusiasm. E1 and E2 are similar terms, appearing together in the same branch of the dendrogram, and part of a cluster that includes Tr2 and KF1. As mentioned before, this cluster is related to the GEMS "Transcendence" factor and therefore no evidence was found to create a new separate subscale. The term E1 was kept (as explained before), and the terms E2 was removed from the list of terms. E3 emerged as part of a new, not hypothesized cluster and will be discussed later.

Anger. The three terms tested for this subscale form an independent cluster, together with T4. Therefore, there is clear evidence of an "Anger" subscale. They are also closely related to the "Tension" subscale.

Section C
GEMIAC: Intensity Checklist

Instructions: In the table below, you will find 14 classes (or families) of feelings. Each class is described by two terms depicting similar feelings. Music often elicits several of these classes of feelings. Your task is to rate the intensity with which you experienced each of the classes of feelings (which can be the specific feelings depicted by one or both of the terms, or similar to these) while listening to a particular piece of music. Please indicate how the piece of music you just listened to made **you** feel (e.g., this music made me feel melancholic/sad). Do not describe the music itself (e.g., this music is melancholic/sad) or what the music may be expressive of (e.g., this music expresses melancholy/sadness). Keep in mind that a piece of music can be melancholic/sad or can sound melancholic/sad without making you feel the same way.

Please rate the intensity with which you experienced each of the following classes of feelings while listening to the piece of music on a scale ranging from 1 (not at all) to 5 (very much).

Not at all	*Somewhat*	*Moderately*	*Quite a lot*	*Very much*
1	*2*	*3*	*4*	*5*
filled with wonder, amazed	1	2	3	4	5
moved, touched	1	2	3	4	5
enchanted, in awe	1	2	3	4	5
inspired, enthusiastic	1	2	3	4	5
energetic, lively	1	2	3	4	5
joyful, wanting to dance	1	2	3	4	5
powerful, strong	1	2	3	4	5
full of tenderness, warmhearted	1	2	3	4	5
relaxed, peaceful	1	2	3	4	5
melancholic, sad	1	2	3	4	5
nostalgic, sentimental	1	2	3	4	5
indifferent, bored	1	2	3	4	5
tense, uneasy	1	2	3	4	5
agitated, aggressive	1	2	3	4	5

Section D
GEMIAC: Frequency Checklist

Instructions: In the table below, you will find 14 classes (or families) of feelings. Each class is described by two terms depicting similar feelings. Music often elicits several of these classes of feelings. Your task is to rate the frequency with which you experienced each of the classes of feelings (which can be the specific feelings depicted by one or both of the terms, or similar to these) while listening to a particular piece of music. Please indicate how the piece of music you just listened to made **you** feel (e.g., this music made me feel melancholic/sad). Do not describe the music itself (e.g., this music is melancholic/sad) or what the music may be expressive of (e.g., this music expresses melancholy/sadness). Keep in mind that a piece of music can be melancholic/sad or can sound melancholic/sad without making you feel the same way.

Please rate the frequency with which you felt each of the following classes of feelings during the work of music on a scale ranging from 1 (never) to 5 (a lot):

Never *1*	*Rarely* *2*	*Sometimes* *3*	*Frequently* *4*	*A lot* *5*

filled with wonder, amazed	1	2	3	4	5
moved, touched	1	2	3	4	5
enchanted, in awe	1	2	3	4	5
inspired, enthusiastic	1	2	3	4	5
energetic, lively	1	2	3	4	5
joyful, wanting to dance	1	2	3	4	5
powerful, strong	1	2	3	4	5
full of tenderness, warmhearted	1	2	3	4	5
relaxed, peaceful	1	2	3	4	5
melancholic, sad	1	2	3	4	5
nostalgic, sentimental	1	2	3	4	5
indifferent, bored	1	2	3	4	5
tense, uneasy	1	2	3	4	5
agitated, aggressive	1	2	3	4	5